

☐ Gr. 1, J.-P. Haslinger, MSc

☒ Gr. 2, DI (FH) G. Horn, MSc

 Name Smola, Thomas

 Aufwand in h 18

Punkte _____ Kurzzeichen Tutor / Übungsleiter _____ / _____

1. Mini Regex
(2 + 4 + 5 + 5 + 8 Punkte)

Gesucht ist eine stark vereinfachte Regular Expressions Implementierung, welche Zeichenkettenvergleiche auf Pascal-Zeichenketten durchführt.

Reguläre Ausdrücke sind spezielle Suchmuster, die neben dem zu findenden Text auch spezielle Metazeichen enthalten. Ein solches Metazeichen ist z. B. der Punkt (.), welcher für exakt ein beliebiges Zeichen in der Zeichenkette steht. Weitere Metazeichen sind auch die eckigen Klammern ([und]), zwischen denen im Suchmuster beliebig viele Zeichen stehen können, von denen dann eines auf das nächste Zeichen im zu durchsuchenden Text passen muss. Das Metazeichen ^ bedeutet Negation des nachfolgenden Elements im Suchmuster. (Siehe auch Beispiele unten!)

Weitere Informationen zum Thema „reguläre Ausdrücke“ finden Sie im Internet (z. B. unter <http://www.regular-expressions.info>).

a) Beginnen Sie mit einem klassischen „Brute-Force“-Algorithmus. Entwickeln Sie eine Funktion, die für einen Text und ein Muster ermittelt, an welcher Stelle das Muster zum ersten Mal im Text auftritt.

b) Erweitern Sie diese Implementierung nun so, dass auch das Metazeichen Punkt (.) so wie oben beschrieben interpretiert wird, wenn es im Suchmuster vorkommt.

<u>Zeichenkette</u>	<u>Suchmuster</u>	<u>Position der Übereinstimmung</u>
Hagenbrg	b.rg	0
Hagenberg	b.rg	6
Hagenburg	b.rg	6
Hagenberg	b..rg	0

c) Weiters soll nun auch die eckigen Klammern unterstützt werden.

<u>Zeichenkette</u>	<u>Suchmuster</u>	<u>Position der Übereinstimmung</u>
Hagenberg	b[eu]rg	6
Hagenburg	b[eu]rg	6
Hagenborg	b[aeiu]rg	0
Hagenbeurg	b[eu]rg	0

d) Um die eckigen Klammern zu erweitern, bauen Sie Bereiche (markiert durch -) ein, die es ermöglichen noch flexibler zu suchen, z.B. um die Suche nach Ziffern oder Buchstaben zu ermöglichen.

<u>Zeichenkette</u>	<u>Suchmuster</u>	<u>Position der Übereinstimmung</u>
Hagenberg	genbe[a-z]g	3
Hagenborg7	borg[0-9]	6
Hagenbeurg	[A-Z]agen[c-f]	0

e) Bauen Sie nun auch noch die Möglichkeit zur Negation in Ihre Lösung ein. Dabei sollen einerseits einfache Zeichen, aber auch Zeichengruppen und Bereiche negiert werden können (siehe Beispiele).

<u>Zeichenkette</u>	<u>Suchmuster</u>	<u>Position der Übereinstimmung</u>
Hagenbergburg	b^erg	10
Hagenbergburg	b^urg	6
Hagenbergburg	b^[eu]rg	0
Hagenborg	b^[e-k]rg	6